Open source cheminformatics in KNIME with the RDKit: update

Gregory Landrum, Ph.D.
NIBR Informatics
Novartis Institutes for BioMedical Research, Basel
8th KNIME Users Group Meeting
Berlin, 26 February 2015
RDKit: What is it?

- Open-source C++ toolkit for cheminformatics
- Wrappers for Python (2.x, 3.x), Java, C#
- Functionality:
  - 2D and 3D molecular operations
  - Descriptor generation for machine learning
  - Database cartridge for substructure and similarity searching
  - Knime nodes
  - IPython integration
  - Supports Mac/Windows/Linux
- BSD license
- http://www.rdkit.org
What is this all about?

Exact same algorithms/implementations accessible from many different endpoints
Knime integration

- *Open-source* RDKit-based nodes for KNIME providing cheminformatics functionality
- Trusted nodes distributed from KNIME community site

- Work in progress: more nodes being added (new wizard makes it easy)
What's there?
What's there?

RDKit
- Converters
  - RDKit From Molecule
  - RDKit To Molecule
  - RDKit From InChI
  - RDKit To InChI
  - RDKit From IUPAC
  - RDKit Canon SMILES
- Modifiers
  - RDKit Add Hs
  - RDKit Remove Hs
  - RDKit Aromatizer
  - RDKit Kekulizer
  - RDKit Salt Stripper
- Calculators
  - RDKit Descriptor Calculation
  - RDKit Calculate Charges
- Geometry
  - RDKit Generate Coords
  - RDKit Optimize Geometry
  - RDKit Add Conformers
  - RDKit Open 3D Alignment
  - RDKit RMSD Filter
- Fingerprints
  - RDKit Fingerprint
  - RDKit Count-Based Fingerprint
  - RDKit Fingerprint Reader
  - RDKit Fingerprint Writer
  - RDKit Diversity Picker
- Fragments
  - RDKit Molecule Fragmenter
  - RDKit Find Murcko Scaffolds
- Searching
  - RDKit MCS
  - RDKit Substructure Filter
  - RDKit Molecule Substructure Filter
  - RDKit Functional Group Filter
  - RDKit Substructure Counter
- Reactions
  - RDKit One Component Reaction
  - RDKit Two Component Reaction
  - RDKit Chemical Transformation
- Viewing
  - RDKit Interactive Table
  - RDKit SMILES Headers
  - RDKit Highlighting Atoms
- Experimental
  - RDKit R Group Decomposition
  - RDKit Structure Normalizer
RDKit support in the new python integration

- RDKit molecules are directly available
- Fingerprints easily accessible
- Full capabilities of the RDKit now available in KNIME
RDKit Maximum Common Substructure (MCS)

- Very flexible MCS implementation
- Includes partial MCS and generic atom/bond types
- Integration with “Group by” node
RDKit Structure Normalizer

- Recognizes and cleans up frequent input problems
- Flags molecules with drawing problems/ambiguities
- Built on Avalon toolkit functionality

[Images of molecular structures with comments on stereochemistry and molecular fragments.]
Wrapping up

- **Acknowledgements:**
  - Manuel Schwarze (NIBR)
  - Bernhard Rohde (NIBR)
  - Thorsten Meinl (Knime)
  - Bernd Wiswedel (Knime)
  - Simon Richards (Lilly)
  - RDKit Community

*Want more? Come to the RDKit Community Nodes Workshop on Friday!*